

# Community Embeddings for Friend Suggestions

Anton I. N. Begehr

*Graduate School of Business*

*National Research University Higher School of Economics*

Moscow, Russia

a.begehr@fu-berlin.de

Prof. Dr. Petr Panfilov

*Graduate School of Business*

*National Research University Higher School of Economics*

Moscow, Russia

ppanfilov@hse.ru

**Abstract**—Graphs, such as social networks, emerge naturally from various real-world situations. Recently, graph embedding methods have gained traction in data science research. Recommender systems are used in a wide range of business applications and are essential for online, e-business models to survive and thrive in the contemporary market. Using graph embeddings for recommendation tasks, have the possibility of improving upon recommender systems, because of data compression, their feature vector format, and sub-quadratic time complexity. Graph and community embeddings generated with ComE BGMM+VI are used to build a recommender system for friend suggestions. ComE BGMM+VI is an alteration of the community embeddings algorithm ComE. ComE BGMM+VI applies a Bayesian Gaussian mixture model and variational inference for community embedding and detection. Recommendations are evaluated by the top- $N$  hit-rate over users with at least 50 friends. A friend suggestions recommender system with a top-10 leave-one-out hit-rate of 43.6% and run-time optimized 32.9% is presented.

**Index Terms**—graph, embedding, community embedding, ComE, recommendations, friend suggestions

## I. INTRODUCTION

Graphs, such as social networks, knowledge graphs, content-rating graphs, and communication networks, emerge naturally from various real-world situations. Analyzing these graphs leads to findings and understanding of the underlying structures, coherences, and dependencies. Recently, methods for embedding graph's nodes into lower-dimensional Euclidean spaces, called graph embeddings, have gained traction in multiple areas of data science research [6].

Due to the rapid growth of the internet and data accumulation, recommender systems are essential for e-business and online business models to survive and thrive in the contemporary market [13]. Modern recommender systems need to take into account the huge amounts of user data generated at all times in big data systems around the world and improve recommendations instead of failing under the thrust of big data overload.

Utilizing graph embeddings for recommendation tasks, has recently gained research traction [11, 12, 7, 15]. The advantages of graph embeddings include data compression and the Euclidean feature vector format [5]. Given these advantages and provided competitive results, graph embeddings have the possibility of greatly improving upon graph-based use-cases like recommender systems.

Community Embeddings, in addition to embedding a graph's nodes through first- and second-order proximity, also

preserve higher-order proximity by embedding clusters present in the graph data. The graph and community embedding algorithm ComE aims to preserve first-, second- and higher-order proximity by embedding a graph's nodes and communities [3].

This work specifically examines community embeddings for friend suggestion recommender systems and evaluates recommendations on social network graph data for the use-case of friend suggestions. Graph and community embeddings generated with ComE BGMM+VI are used to develop a friend suggestions recommender system based on the shortest distances between nodes in the embedding. Recommendations are evaluated by the top- $N$  recommendations hit-rate of test edges. A friend suggestions recommender system with a top-10 leave-one-out hit-rate of 43.6% and run-time optimized 32.9% is presented.

## II. FRIEND SUGGESTIONS

Recommender Systems are eagerly researched in academia and widely deployed in real-world business applications. Most contemporary technology companies heavily rely on recommender systems to drive usage of their services and consumption of their content. Users, in turn, rely on recommender systems to find what they want and need and save searching time. State-of-the-art recommender systems provide a competitive advantage desperately needed by online services. Companies heavily relying on recommender systems include YouTube, Amazon, Netflix, and many more [14].

Recommender systems are built on top of user-item interaction. Friend suggestions are a simpler type of recommender system. For friend suggestions, no distinction is made between users and items. The entity user is both the subject and object of recommendation. This results in a simple data model, which can be used for structural friend suggestions, as shown in Fig. 1:

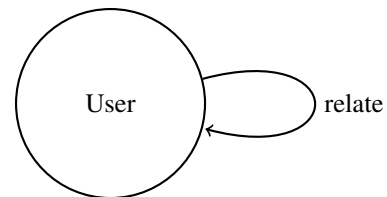


Fig. 1. Minimal data model supporting friend suggestions.

The two main methods used for recommender systems are collaborative filtering and content-based filtering. Collaborative filtering is based on the assumption that users like items similar to other items they like and items that are liked by other users with similar tastes [8, 16]. Content-based filtering considers user and item attributes, instead of solely interactions [14, 9].

In this paper, a method of generating friend suggestions based solely on graph data structure, and not on the node or edge attributes, is presented. The graph of users and friendships is embedded to a lower-dimensional space with the community embedding algorithm ComE BGMM+VI [3, 2, 1]. The approach can be considered a variant of collaborative filtering with the euclidean distance of user embeddings and user community membership as measures for user similarity. The proposed advantage of using such an approach is that community embeddings optimize first-, second-, and higher-order proximity between users in the node embedding.

#### A. Evaluation

To determine the effectivity of friend suggestions, the generated friend suggestions must be properly validated. In this paper, the top- $N$  approach to evaluating recommender systems with hit-rate as the evaluation metric is chosen.

A user's top- $N$  recommendations is a list of  $N$  items to be recommended to a specific user. To evaluate recommender systems with the top- $N$  and hit-rate approach, initially, the dataset is split into train and test data. For each testing user, one relation is left out, according to the leave-one-out method and the model is trained on the remaining training dataset. Once the model is trained, a list of top- $N$  recommendations is generated for each testing user. If the item corresponding to the user is in the user's top- $N$  list, a hit is counted, otherwise, a miss is counted. The hit-rate is defined as the total number of hits divided by the number of testing users.

Utilizing the top- $N$  approach with hit-rate to evaluate recommender systems is advantageous to evaluating recommender systems by a link prediction approach since the top- $N$  approach is more realistic in comparison to actual recommender system use-cases. When you open Netflix, Amazon, YouTube, or Facebook friend suggestions, one or multiple top- $N$  recommendation lists are generated and displayed. If you click on an item and buy, watch, or befriend, that is considered a hit, otherwise, a miss.

The hit-rate metric on top- $N$  recommendations provides a realistic option of evaluating recommender systems [4, 10, 17].

### III. ALGORITHM

The proposed algorithm for generating friend suggestions using community embeddings is detailed and evaluated on time complexity. The initially quadratic runtime of generating friend suggestions for all users is then reduced by a factor of  $K$  by utilizing a node's community membership. The two resulting algorithms are presented.

### REFERENCES

- [1] Anton Begehr. *abegehr/ComE\_BGMM*. 2020. URL: [https://github.com/abegehr/ComE\\_BGMM](https://github.com/abegehr/ComE_BGMM) (visited on 05/09/2020).
- [2] S. Cavallari, E. Cambria, H. Cai, K. C. Chang, and V. W. Zheng. "Embedding Both Finite and Infinite Communities on Graphs [Application Notes]". In: *IEEE Computational Intelligence Magazine* 14.3 (2019), pp. 39–50.
- [3] Sandro Cavallari, Vincent W. Zheng, Hongyun Cai, Kevin Chen-Chuan Chang, and Erik Cambria. "Learning Community Embedding with Community Detection and Node Embedding on Graphs". In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. CIKM '17. Singapore, Singapore: Association for Computing Machinery, 2017, 377–386. ISBN: 9781450349185. DOI: 10.1145/3132847.3132925. URL: <https://doi.org/10.1145/3132847.3132925>.
- [4] Paolo Cremonesi, Yehuda Koren, and Roberto Turrin. "Performance of Recommender Algorithms on Top-N Recommendation Tasks". In: Jan. 2010, pp. 39–46. DOI: 10.1145/1864708.1864721.
- [5] Primož Godec. *Graph Embeddings — The Summary*. 2018. URL: <https://towardsdatascience.com/graph-embeddings-the-summary-cc6075aba007> (visited on 08/27/2020).
- [6] Palash Goyal and Emilio Ferrara. "Graph Embedding Techniques, Applications, and Performance: A Survey". In: *Knowledge-Based Systems* 151 (July 2018), 78–94. ISSN: 0950-7051. DOI: 10.1016/j.knosys.2018.03.022. URL: <http://dx.doi.org/10.1016/j.knosys.2018.03.022>.
- [7] László Grad-Gyenge, Attila Kiss, and Peter Filzmoser. "Graph Embedding Based Recommendation Techniques on the Knowledge Graph". In: July 2017, pp. 354–359. DOI: 10.1145/3099023.3099096.
- [8] Prince Grover. "Various Implementations of Collaborative Filtering". In: *Towards Data Science* (Dec. 28, 2017). URL: <https://towardsdatascience.com/various-implementations-of-collaborative-filtering-100385c6dfe0> (visited on 08/08/2020).
- [9] Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. "Content-based Recommender Systems: State of the Art and Trends". In: Jan. 2011, pp. 73–105. DOI: 10.1007/978-0-387-85820-3\_3.
- [10] Enrico Palumbo, Giuseppe Rizzo, and Raphaël Troncy. "Entity2rec: Learning User-Item Relatedness from Knowledge Graphs for Top-N Item Recommendation". In: *Proceedings of the Eleventh ACM Conference on Recommender Systems*. RecSys '17. New York, NY, USA: Association for Computing Machinery, 2017, 32–36. ISBN: 9781450346528. DOI: 10.1145/3109859.3109889. URL: <https://doi.org/10.1145/3109859.3109889>.
- [11] Enrico Palumbo, Giuseppe Rizzo, Raphaël Troncy, Elena Baralis, Michele Osella, and Enrico Ferro.

- “An Empirical Comparison of Knowledge Graph Embeddings for Item Recommendation”. In: *DL4KGS@ESWC*. 2018.
- [12] Enrico Palumbo, Giuseppe Rizzo, Raphaël Troncy, Elena Baralis, Michele Osella, and Enrico Ferro. “Knowledge Graph Embeddings with node2vec for Item Recommendation”. In: *ESWC*. 2018.
- [13] Nikolaos Polatidis and Christos K. Georgiadis. “Recommender Systems: The Importance of Personalization in E-Business Environments”. In: *IJEEI* 4 (2013), pp. 32–46.
- [14] Baptiste Rocca. “Introduction to Recommender Systems”. In: *Towards Data Science* (June 3, 2019). URL: <https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada> (visited on 08/08/2020).
- [15] Vishwas Sathish, Tanya Mehrotra, Simran Dhinwa, and Bhaskarjyoti Das. “Graph Embedding Based Hybrid Social Recommendation System”. In: *ArXiv abs/1908.09454* (2019).
- [16] Xiaoyuan Su and Taghi M. Khoshgoftaar. “A Survey of Collaborative Filtering Techniques”. In: *Adv. Artif. Intell.* 2009 (2009), 421425:1–421425:19.
- [17] Z. Zhao, Ming Zhu, Y. Sheng, and Jinlin Wang. “A Top-N-Balanced Sequential Recommendation Based on Recurrent Network”. In: *IEICE Trans. Inf. Syst.* 102-D (2019), pp. 737–744.